DOCUMENT RESUME

ED 361 407 TM 020 534

AUTHOR Davidson, Fred
TITLE The Testing of English as a Second/Foreign Language in the Criterion-Referenced Era.
PUB DATE Nov 92
NOTE 19p.; Paper presented at the Annual Meeting of the National Council of Teachers of English (Louisville, KY, November 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Criterion Referenced Tests; Educational Assessment; Elementary Secondary Education; *English (Second Language); Language Proficiency; *Language Tests; Limited English Speaking; Models; Norm Referenced Tests; *Second Language Learning; *Student Evaluation; *Test Construction; Test Items
IDENTIFIERS Criterion Referenced Language Test Development

ABSTRACT

In the assessment of second/foreign language proficiency, we are entering the era of criterion-referenced assessment as language learning is being recognized as an integrative, multifaceted construct. Norm-referenced measurement (NRM) is compared with criterion-referenced measurement (CRM). CRM is characterized by attention to skill, whereas NRM focuses on student rank. The evolution of some modern multi-componential language ability models is traced, starting with that of Canale and Swain (1980). CRM, with its greater focus on skills, should provide a better perspective from which to measure such a wide array of skills than NRM. One process for measuring skills is criterion-referenced language test development (CRLTD). CRLTD is characterized by flexibility. Participants iterate between the test planning and test item/task writing, cycling between test specification and product. CRLTD is a bottom-up group-based test development process. Although time constraints do not permit a thorough exploration of the technique in this paper, audience members are encouraged to try it to improve skills-based testing in the modern era of language teaching. Six figures illustrate the discussion. (Contains 8 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 361 407

(M)

The Testing of English as a Second/Foreign Language in the Criterion-Referenced Era.

by Fred Davidson
University of Illinois at Urbana-Champaign

[Paper given at the NCTE Convention, Louisville, KY, 20 NOV 92. For further information, please contact the author via the address given at the end of the paper.]

[Handout: Abstract, references and address (on one side), Figure 6 below on the other]

ABSTRACT

This paper begins by comparing norm-referenced measurement (NRM) with criterion-referenced measurement (CRM). CRM is characterized by attention to skill whereas NRM focuses on student rank. Next, the paper goes through the evolution of some modern multi-componential language ability models, starting with Canale and Swain (1980). CRM, with its greater focus on skill, should be a better perspective to measure such a wide array of skills than NRM. One process to do so is CRLTD: criterion-referenced language test development. Time does not permit thorough experience with CRLTD today, but audience members are encouraged to try it at their educational institutions, to better effect skills-based testing in the modern, complex language teaching era.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
FRED DAVIDSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM020534

BEST COPY AVAILABLE

1. CRM vs. NRM:

There is an undeniable need to assess in educational settings. This need derives, largely, from the need to make decisions about people. We need to decide about placement into course sequences, about aptitude to learn material, about achievement of material once taught, and about diagnosis when something seems to have gone wrong. All these needs seem to breed tests.

A tradition of testing has emerged over the last hundred years. This tradition says that the best way to assess in education is to rank students along some sort of trait continuum. To assess height, you can line the kids up and see who is tallest, who is next tallest, and so on. That works fine for height. If you want to know who is tallest in your class, line them up and compare.

But language ability is not like height. Let's examine a more challenging problem: assessing the English proficiency of a language minority student in some hypothetical K-12 setting. In the USA, generally, to be labeled a 'language minority student', the student must fulfill two criteria: (1) she or he comes from a home environment where English is NOT the predominant language, AND (2) she or he lacks sufficient command of English to be able to compete with her or his grade/age peers. These two criteria: the 'home language' and the 'proficiency' are reflected in plenty of state and national laws, for example, Article 14c of the School code of my home state, Illinois.

Let's focus only at the second of those two needs: determining if the student has sufficient command of English to

compete with her or his grade/age peers.

From the tradition I just mentioned, you'd have to be able to line the kids up and see who is 'tallest' -- who has the best command of English. If the language minority student wound up at the 'short end', then some sort of English support might be necessary. But the problem here is that the particular group you are investigating -- that mix of kids -- is serving as a 'norm'. You are fixing a decision about the language minority student relative to that norm, and the norm may be somehow unique or particular to that group. This is known as norm-referenced measurement; the decision about our language minority student is based on her or his rank among grade/age peers.

Missing in this formula is some sort of attention to what it means to command English like the peer group. We don't get any absolute understanding of what English skills the student does or does not have. What does proficiency mean? Does it mean answering a bunch of discrete multiple-choice grammar questions? Does it mean the ability to conduct a role-play with the teacher in English? Does it mean the sensitivity to switch from one register to another, as when speaking to a beloved pet versus speaking to the school principal? Well-developed norm-referenced measures do pay attention to content, but so long as the norm-referenced test instruments consistently rank students and compare well to other norm-referenced tests, content is secondary. Stability of results and predictability of decisions is more important under norm-referencing than careful attention to language skills.

This odd state of affairs is changing, and as my title

suggests, I believe it has already changed. We are now more interested in content than rank. We are in an era where the result of the test is anchored, or 'referenced' to some identifiable task or set of tasks. In second/foreign language assessment, we are in the era of Criterion-Referenced Measurement. I believe this to be true because there have been vast changes in our perspectives about language ability. We no longer see language competence as a monolithic single trait, best assessed by an aggregate score on a collection of discrete test questions. We no longer view language learning as the acquisition of zillions of little bits. We see it as an integrative, multifaceted construct. And that demands a change in our perspective on language testing as well.

Some very important developments in second/foreign language theory had lots to do with this. Let me outline one major influence: the post-Canale and Swain 'movement'.

2. Attention to plethora of skills in the post-Canale and Swain era.

An excellent reference to the nature of language teaching and second/foreign language learning is H. Douglas Brown's 1987 Principles of Language Learning and Teaching, published by Prentice-Hall. It is remarkably readable, and it is a frequent text in second language acquisition courses. In Chapter 10, Brown discusses the concept of 'communicative competence'. Communicative competence is the umbrella term for the wide range of skills involved in second/foreign language learning. I cannot really summarize communicative competence as well as Brown does, so I am going to allow his words to speak here. Brown states:

[BEGIN QUOTE]

Seminal work on defining communicative competence was carried out by Michael Canale and Merrill Swain (1980), now the reference point for virtually all discussions of communicative competence vis-a-vis second language teaching. In Canale and Swain's (1980) [ref. ohp/fig. 1] and later Canale's (1983) [ref. ohp/fig. 2] definition, four different components, or subcategories, make up the construct of communicative competence. The first two subcategories reflect the use of the linguistic system itself. [ref. ohp/fig. 3 -- Brown is making a slight adjustment to the original Canale and Swain model] Grammatical competence is that aspect of communicative competence that encompasses 'knowledge of lexical items and of rules of morphology, syntax, sentence grammar semantics, and phonology' (Canale and Swain, 1980:29). It is the competence that we associate with mastering the linguistic code of a language. ... The second subcategory is discourse competence, the complement of grammatical competence in many ways. It is the ability we have to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances. Discourse means everything from simple spoken conversation to length written texts (articles, books, and the like). While grammatical competence focuses on sentence-level grammar, discourse competence is concerned with intersentential relationships.

The last two subcategories define the more functional

aspects of communication. Sociolinguistic competence is the knowledge of the sociocultural rules of language and of discourse. This type of competence "requires an understanding of the social context in which language is used: the roles of the participants, the information they share, and the function of the interaction. ... The fourth category is strategic competence, a construct that is exceedingly complex. Canale and Swain (1980: 30) described strategic competence as 'the verbal and nonverbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or due to insufficient competence.'

[END QUOTE]

From the original Canale and Swain 1980 paper, what we have, then, is a model of language ability that looks like Figure 1 [ref: ohp/fig. 1]: communicative competence is separated into three competencies: grammatical competence, sociolinguistic competence, and strategic competence. I should clarify that 'grammatical competence' is used to refer not only to sentence-level grammar rules, but to all the 'systems' of language: grammar, discrete vocabulary rules, morphology, phonology, and so on. Then as shown in Figure 2 [ref: ohp/fig. 2], Canale's 1983 paper adds 'discourse competence'. My impression is that these four are widely accepted in the language teaching field. Brown adds a slight twist in that he separates the four into two groups, the linguistic system and the functional aspects of communication [ref. ohp/fig. 3].

The four aspects of language ability each define a unique

domain of skill. Each does something separate, yet each is related to the other. For example, the ability to use sentence-level grammar is related to discourse command. Or for example, the ability to plan an utterance, especially if one is not yet fully proficient, is related to sociolinguistic rules of formality.

Others have picked up the theme of the post-Canale and Swain movement. That movement is characterized by a firm belief that language competence is multi-componential. Our mandate is to improve the language ability in our students, and that ability is a complex, multi-faceted beast indeed. Bachman (1990) evolves this model further; he elaborates his model of communicative language ability but adds a whole chapter on the complexity of modeling test method -- the TYPE of test question as opposed to WHAT it measures [ref. ohp/fig. 4]. Time does not permit, today, thorough investigation of these later complex models.

What is significant about the post-Canale and Swain vision of language ability? Why is it important in the criterion-referenced era of EFL/ESL testing?

I contend that a multifaceted understanding of language ability is a major progressive step in language teaching and testing. Prior to the work of Canale and Swain, and the critical work of Sandra Savignon (e.g. Savignon 1983), language tests were pretty much norm-referenced and highly discrete. They were monolithic aggregates of many small language skills, most typically highly isolated grammar or reading and vocabulary, which viewed language ability as a single trait. These skills, I contend, are largely from the 'Grammatical' competence component

of the Canale and Swain perspective. [ref. ohp/fig. 2] These tests were like that because they were easy to develop. Norm-referencing worked well: write a bunch of items -- a bunch more than you need (like 5:1) -- and save only those which appear to work well statistically. Tailor made for a monolithic approach to language ability, e.g. grammatical linguistic competence alone because you could write hundreds of questions on discrete grammatical and vocabulary points and save only those which displayed good statistical quality after pretesting.

Yet we hope [return to ohp/fig. 2] that language also includes integrative competency in discourse, sociolinguistic rules, and strategic planning. I maintain that in order to test those we have to have a criterion-referenced view of language testing. It is necessary to formulate our curricula and theory with a clear understanding of the complexity of our charge, and blind norm-referenced measurement does not measure up. We must pay attention to skill, not only rank.

3. The two come together: CRLTD.

I'd like to sketch a procedure that can address the need for a better attention to the multiplicity of skills in current language teaching: Criterion-referenced language test development, or CRLTD. CRLTD is characterized by flexibility. Test development is seen as a series of steps, each connected to the other with a feedback channel. A good CRLTD test is never finished; it is always getting feedback from other steps in the process. Figure 5 shows a schematic of this development process [ref. ohp/fig. 5]. No step is isolated. Each is part of an

ongoing, fluid, integrated whole.

As our job has become more multifaceted, so too has our test development. Brian Lynch and I propose (Davidson and Lynch, forthcoming) that anyone can 'sense' the flux and fluidity of Criterion-referencing in the modern era, by conducting a CRLTD workshop. Figure 6 shows the basic steps of a CRLTD workshop [ref. ohp/fig. 6 -- Figure 6 is on the back of your handout]. The key element in this figure is that the participants iterate between the test planning and test item/task writing: they cycle between 'spec' (specification) and product. The spec writers communicate with the item writers, and gradually the proper assessment technique emerges, given the grouped understanding of all participants.

One key feature of CRLTD is that it is a bottom-up, group based consensus test development process. The interpretation of the 'mandate' (step 3 in figure 6) is open to all involved. That mandate may involve attention to the complexity of current language ability models, such as I have shown. As the group works on its criterion-referenced test, it is free to interpret and re-interpret the meaning of language ability models and fit them to the local needs. This is locally appropriate technology, in which the test is tuned to an institution's own goals and perspectives.

Key to doing this is the role of the Criterion Referenced Specification, or plan. I don't have much time today to go into the nature of a spec. Given more time, I'd hold a workshop here and let you pick a mandate and experience all of Figure 6. I would like to note that a spec is central to the workshop

outlined in Figure 6. Most any planning rubric or outline would do -- alternatively, you can use the one that Brian and I propose in our paper: the style developed by Popham (1978, 1981) in the 1960s and 70s. The principle is the same: the workshop involves communication between the test planner or 'specifier' (step 4 in Figure 6) and the test item or task 'writer' (step 5 in Figure 6). The more times you repeat this process the better these people are able to communicate, and the better they can communicate the better they can interpret the mandate -- even if it is a highly complex multi-faceted language ability model.

4. Conclusion: The Priesthood and you.

Norm-referenced measurement was -- and still is -- run by a statistically ordained priesthood. To practice it, you have to go to 'seminary': you have to get a solid Ph.D. in educational measurement so that you can speak the Latin of statistics. There is nothing amiss with this metaphor, and if I can switch gears a bit, I do tend to agree with Anne Frank: 'People are basically good at heart.' Certainly Priests are. I am not saying that the Norm-referenced establishment is anti-education or anti-learning. Nor am I advocating that we throw out large norm-referenced tests like the TOEFL, the SLEP, the S.A.T., the A.C.T. or others. I am advocating that we supplement such tests with criterion-referenced measures which pay attention to skill as well as rank. And I am offering a means to do so: iterative CRLTD.

One benefit of CRLTD should be heightened content validity. Content validity, in this case, is the link between testing and teaching. A test is content-valid if it accurately and

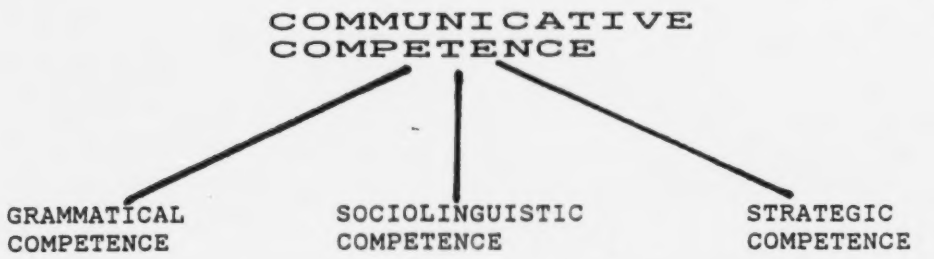
thoroughly reflects the content of instruction in a particular setting. In our example above, the placement exam to decide about a new language minority student should be 'content-valid' to forthcoming instruction. It should reflect the kinds of skills a student is expected to learn during ESL/EFL instruction at that institution. Through CRLTD you can evolve this content validity link.

You can try out CRLTD. I have left Figure 6 on the ohp and have provided it on your handout on purpose to let you consider that such a workshop is actually feasible at your setting, perhaps during your next teacher in-service day. Be sure to run the workshop completely, and preferably at least twice, as step 7 in Figure 6 suggests.

Teaching and assessing language minority students is a complex job. Consider again that ostensibly simple placement need I mentioned at the beginning. The complexity of skills and abilities involved there is mind-boggling. Certainly grammatical competence is involved. Certainly, too, are sociolinguistic rules of appropriacy. Certainly also are competences in discourse organization and strategic language planning. Our job is not easy: dealing with language minority students for whom English is a foreign language. Testing is doubly difficult due to the social decisions in which it operates. But criterion-referencing and solid CRLTD allows a voice to people who are not normally heard: the congregation (you) as well as the priests (the psychometricians).

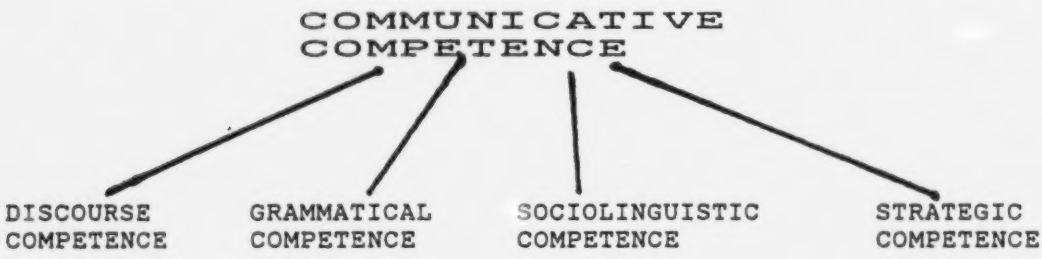
Please, speak up.

Figure 1.



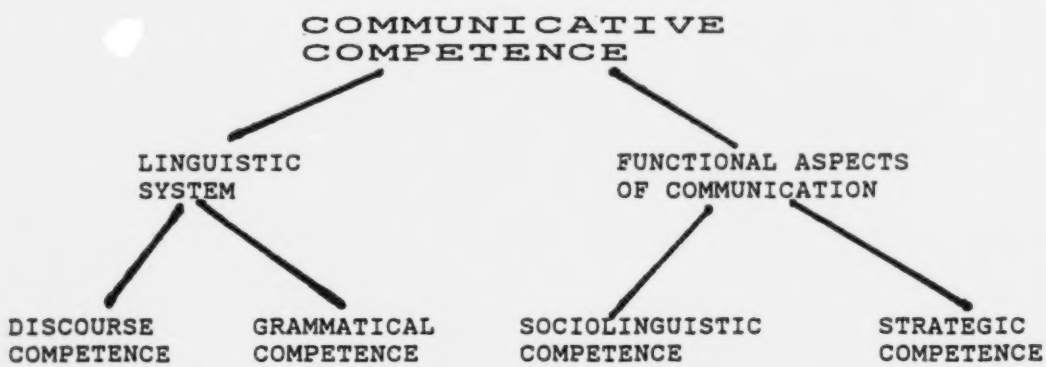
(Canale and Swain, 1980)

Figure 2.



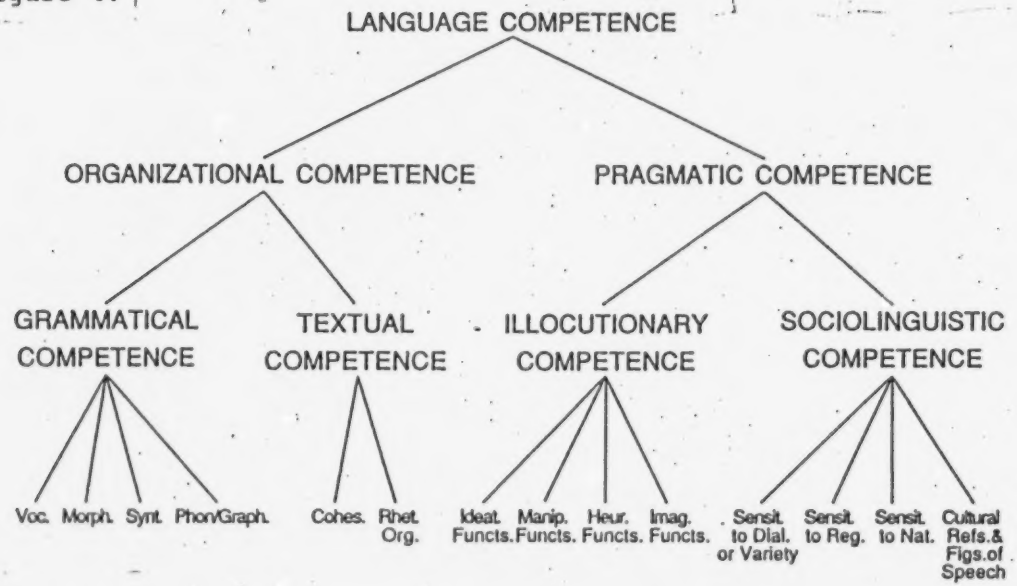
(Canale, 1983)

Figure 3.



(Brown, 1987: 199-200)

Figure 4.

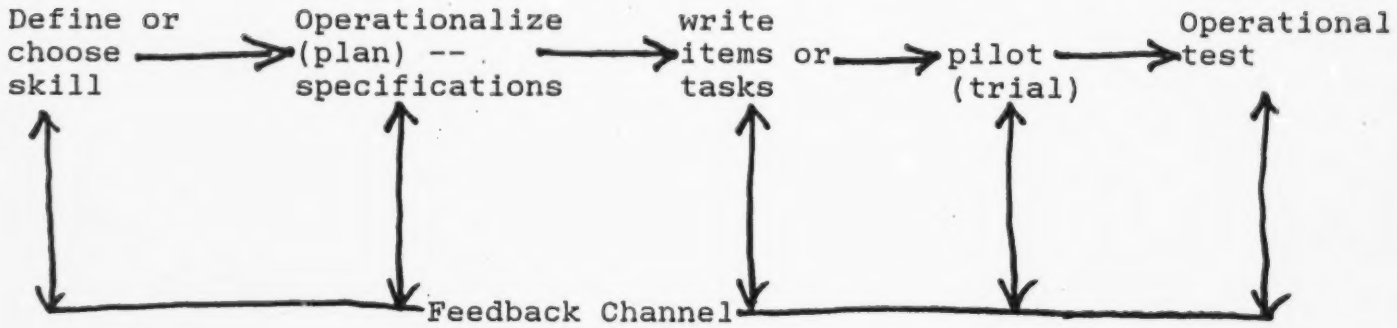


(Bachman, 1990: 87)

Fred Davidson, NCTE 1992, p.14

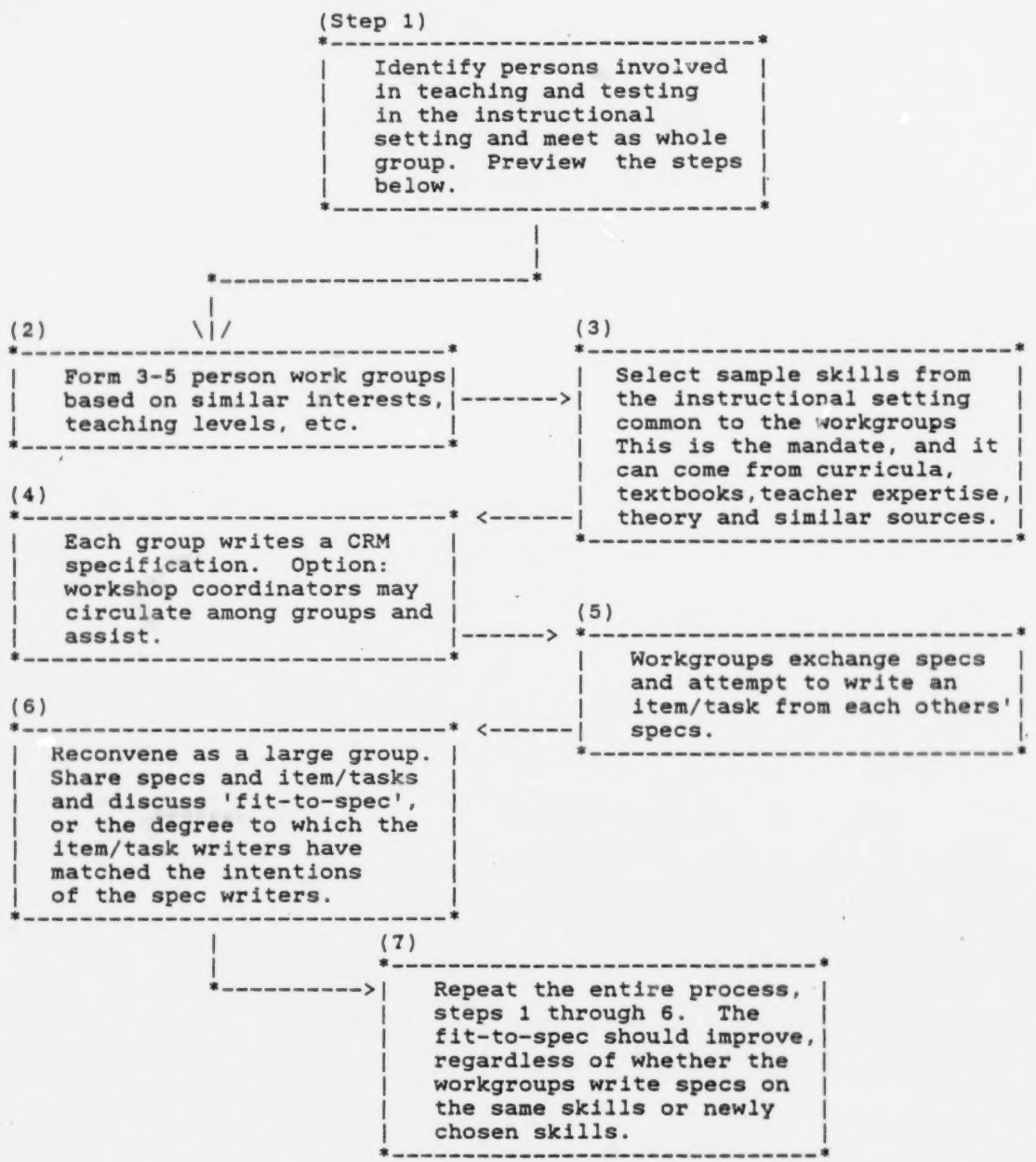
Figure 5.

Steps in Test Development



Fred Davidson, NCTE 1992, p.15.

Figure 6: Steps in a CRLTD Workshop:


(Step 1)
Identify persons involved in teaching and testing in the instructional setting and meet as whole group. Preview the steps below.
(2)
Form 3-5 person work groups based on similar interests, teaching levels, etc.
(3)
Select sample skills from the instructional setting common to the workgroups This is the mandate, and it can come from curricula, textbooks, teacher expertise, theory and similar sources.
(4)
Each group writes a CRM specification. Option: workshop coordinators may circulate among groups and assist.
(5)
Workgroups exchange specs and attempt to write an item/task from each others' specs.
(6)
Reconvene as a large group. Share specs and item/tasks and discuss 'fit-to-spec', or the degree to which the item/task writers have matched the intentions of the spec writers.
(7)
Repeat the entire process, steps 1 through 6. The fit-to-spec should improve, regardless of whether the workgroups write specs on the same skills or newly chosen skills.

(from Davidson and Lynch, forthcoming)

REFERENCES

Bachman, Lyle F. 1990. Fundamental Considerations in Language Testing. Oxford, UK: Oxford University Press.

Brown, H. Douglas. 1987. Principles of Language Learning and Teaching, 2nd. Edition. Englewood Cliffs, NJ: Prentice-Hall.

Canale, Michael. 1983. From communicative competence to communicative language pedagogy. In Jack C. Richards and Richard Schmidt (Eds). Language and Communication. London: Longman.

Canale, Michael and Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics 1:1, pp. 1-47.

Davidson, Fred and Brian Lynch. forthcoming. Criterion-Referenced Language Test Development: A Prolegomenon. To appear in the collected papers from the August 1991 meeting: "Language Testing in the 1990s" held at the University of Jyvasklya, Finland.

Savignon, Sandra. 1983. Communicative Competence: Theory and Classroom Practice. Reading, MA: Addison-Wesley

Popham, W.J. 1978. Criterion-Referenced Measurement. Englewood Cliffs, NJ: Prentice-Hall.

Popham, W.J. 1981. Modern Educational Measurement. Englewood Cliffs, NJ: Prentice-Hall.

==========

Author's address:

Fred Davidson, Assistant Professor
Division of English as an International Language (DEIL)
3070 Foreign Languages Building (FLB)
University of Illinois at Urbana-Champaign (UIUC)
707 South Mathews
Urbana, IL 61801, USA
tel: +217-333-1506
fax: +217-244-3050
computer mail:
Bitnet: davidson@uiucvmd
Internet: davidson@vmd.cso.uiuc.edu
Bitnet passthru: davidson@uiucvmd.bitnet